DOCUMENT RESUME

ED 310 130 TM 013 703

AUTHOR Pardo, Elly; Russell, Joanne Collins
TITLE Standardized Tests at the Early Childhood Level: Do They Tell Us the Truth about Children?
PUB DATE Mar 89
NOTE 45p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, March 27-31, 1989).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150) -- Tests/Evaluation Instruments (160)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Academic Achievement; *Achievement Tests; Children; *Early Childhood Education; *Elementary School Students; Followup Studies; Grade 1; Grade 2; Grouping (Instructional Purposes); High Risk Students; Remedial Reading; *Standardized Tests; *Test Validity; *Writing Evaluation
IDENTIFIERS *Metropolitan Achievement Tests

ABSTRACT

A small follow-up study of three male and five female children, who were at risk of school failure and were enrolled in a model developmental Early Learning Center in the Boston (Massachusetts) Public Schools, is described. The focus was on determining whether a standardized achievement measure used to evaluate student academic success/progress accurately reflected the level of mastery attained by the students. Despite their satisfactory performance in reading/language arts and mathematics, most of the students scored below the 40th percentile when tested at the end of the school year on the Metropolitan Achievement Test (MAT6) in reading and mathematics. The students' second-grade teachers were interviewed to determine their assessment of the students' performance vocabulary, word-recognition, reading comprehension, and writing skill areas. The students' journals were rated by independent judges, and the first-grade achievement of follow-up students on the school department's reading/language arts and writing curriculum-referenced tests were compared with their first-grade achievement on the MAT6 reading test. The MAT6 did not articulate well with other measures of the students' first-grade reading/language arts skills or their second grade reading/language arts school performance. The use of a single test score to classify children for special educational services, particularly in the early childhood years, is questioned. Discrepancies have arisen between the format for classroom learning at the early childhood level and the format for measuring learning. Five data tables are included; and interrater reliability coefficients, the student follow-up survey, transcription rules, samples of transcribed texts, and the writing sample evaluation instrument are appended. (TJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED310130

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

ELLY B. PARDO

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Standardized Tests at the Early Childhood Level: Do They Tell Us the Truth about Children?[1]

Elly Pardo and Joanne Collins Russell

Boston Public Schools

Paper prepared for delivery at the annual meeting of the American Educational Research Association, San Francisco, CA, March 27-31, 1989.

[1] Viewpoints expressed in this document are solely those of the authors. Reprint requests can be directed to: Dr. Elly Pardo, Department of Program Evaluation, Boston Public Schools, 26 Court St., Boston, MA 02108, (617) 726-6200 x5793, or Joanne Collins Russell, Early Learning Center, 50 Beechcroft Street, Brighton, MA 02135, (617) 254-6672.

TM013703

BEST COPY AVAILABLE

HIGHLIGHTS

This paper describes a small, follow-up study of eight children enrolled in a model, developmental Early Learning Center (ELC) in the Boston Public Schools (BPS). During the 1987 - 1988 school year, the children in question were assigned to a first-grade classroom in which learning occurred through direct, hands-on experiences. Despite their average or above average classroom performance in reading/language arts and math, the majority of youngsters scored below the 40th percentile when tested at the end of the school year on the Metropolitan Achievement Test (MAT6) in Reading and Math. Given that the 40th percentile is the point used by the Boston Public Schools to identify students who are potentially *at risk* of academic failure, it was decided to investigate whether the MAT6 had accurately reflected the level of academic attainment and learning potential of the students under study.

The analyses in this document address only the area of reading/ language arts. They point to the mismatch that has occurred between the assessment of student skills by a standardized achievement test and the assessment of those same skills by other indicators of student academic performance.

THEORETICAL BACKGROUND: CHILD LEARNING THEORY

- Children learn best when they acquire information at their own pace through direct, hands-on experiences.
- Children construct their notions of the world by physically interacting with it and not by passively observing or reacting to things and events in their environment.
- Standardized achievement tests evaluate skills through an abstract, decontextualized format that is based entirely upon symbolic reasoning. This format is incompatible with the way young children naturally process information.
- Even though students may be 'test-ready,' that is, they may have mastered good test-taking skills, they may not have developed the theoretical constructs necessary for them to see relationships among ideas or to apply information to a variety of learning environments.
- When imitation is taught at the expense of experiential learning, complex reasoning and problem-solving skills that are critical for higher-order thinking are bypassed.

METHODS

- In a structured telephone interview conducted in February of 1989, second-grade teachers of follow-up pupils were asked to rate the classroom performance of these pupils in skills areas pertaining to vocabulary, word-recognition, reading comprehension, and writing.

- Journals produced by ELC follow-up students in April of 1988 were independently rated by two first-grade teachers from schools located in the same geographical area. The teachers were using instructional techniques in reading/language arts that articulated well with the format of the Metropolitan Achievement Reading Test -- that is, they were teaching primarily through a basal reading series, workbooks, and topic-specific worksheets.

- The first-grade achievement of follow-up students on the school department's Reading/Language Arts and Writing Curriculum-Referenced Tests was compared with their first-grade achievement on the MAT6 Reading test. Teacher judgements of pupil performance in reading, language arts, and writing were also compared for both first and second grades.

FINDINGS

- Outcomes on the Metropolitan Achievement Test in Reading for follow-up pupils show that the instrument did not articulate well with other measures of their first-grade reading/language arts skills level or their second grade reading/language arts school performance.

- The school department's Grade 1 Curriculum-Referenced Test in Reading/Language Arts provided a more realistic picture of the second-grade reading/language arts performance of follow-up students than did the MAT6 in Reading. This is largely because test items on the Reading/Language Arts CRT not only match the content of the curriculum for the BPS, but they also assess student knowledge appropriate for a specific grade level. Items on the Metropolitan Achievement Tests, however, are based upon information in curriculum guides used nationally, and the material on these tests spans several grade levels.

- Second-grade teacher ratings of student reading, language arts, and writing school performance revealed that all the youngsters in question were on grade level in these skills areas. In fact, even though they qualified for and were receiving Chapter 1 (federally-funded) remediation services, some of the youngsters were judged by their teachers to be at the top of their class in literacy skills.

- First-grade teacher ratings of student journals were highly favorable and also bore some resemblance to the ratings of student writing ability by second-grade teachers.

- First-grade CRT Writing scores for observed students generally were consistent with teacher evaluations of their writing ability in both the first and second grade.

IMPLICATIONS OF FINDINGS

- The data in this study suggest that an assessment tool that measures skills taught beyond a child's actual grade level may be categorizing more students as being at educational risk than is actually the case.

- The data also raise questions about the use of a single test score to label or classify children for special educational services, particularly in the early childhood years when the growth patterns of youngsters are uneven.

- Finally, the study draws attention to potential problems that have arisen between the format for classroom learning at the early childhood level and the format for measuring learning achievement. More specifically, the contents of the document set forth a dilemma between current theory on the most educationally-sound practices in early childhood education and the use of a standardized achievement test to successfully capture the impact of those practices.

INTRODUCTION

In this paper we describe a small, pilot study investigating whether a standardized achievement test, which measures discrete items of knowledge for specific content areas, can successfully capture the abilities of young children acquiring skills through an integrated, discovery-oriented approach to learning. We suggest that because standardized measures retrieve information in ways that do not parallel the developmental learning processes of young children, they may be better indicators of good test-taking skills than of actual knowledge mastery. Evidence for our contentions comes from psychological learning theory (Piaget, 1965, 1970), observational data on the classroom of eight, first-grade students enrolled in a developmental education Early Learning Center (ELC), Metropolitan Achievement Test (MAT6) scores in reading for the eight observed students, analyses of student products, and follow-up data on student classroom performance.

Children attending the Early Learning Center in question were participating in a model developmental education program for both the Boston Public Schools and the State of Massachusetts. Consequently, a number of school-department administrators considered it important to conduct follow-up research on ELC first graders who had performed poorly on the MAT6 Reading test but who had demonstrated school academic skills that enabled them to be promoted to the second grade.

For the Boston Public Schools, poor performance on the Metropolitan Achievement Test is indicated by a score that falls at or below the 40th percentile. This percentile rank is highly important within the school department for three reasons: First, it is the point at which students are identified as being at educational risk -- that is, they are labeled as potentially in danger of failing school; second, it is the point which determines pupil eligibility for Chapter 1 remediation services;[1] and third, it is the point which flags a school, its instructional programs, its teachers, and its administrators as not having met the school system's academic standards.

The overarching purpose of a study that contrasts test performance with school achievement is to determine the level of compatibility between a standardized achievement measure used to evaluate student academic success and other indicators of student academic progress. The study described here not only examines this question but also considers a related concern, which is the extent to which a standardized test score can reliably account for a pupil's actual and potential academic attainment.

---

[1] Chapter 1 is a federal entitlement program for low achieving students.

The discussion that follows provides descriptive information on the first-grade learning environment of follow-up pupils as well as comparative data on student test scores and teacher evaluations of pupil academic performance. Analyses are presented within both a qualitative and quantitative framework and address the mismatch that has occurred between student achievement and the method used to assess instructional progress.

THE INSTRUCTIONAL CONTEXT

During the 1987-1988 school year, the students in question attended a first-grade classroom in a full-day developmental education program having before- and after-school daycare services. The philosophical assumption underlying instruction in a developmental model is that children learn best when they acquire information at their own pace through direct, hands-on experiences (NAEYC, 1988). These experiences, often individualized, are thought to allow preschoolers sufficient room for exploration, creativity, and the formation of good critical-thinking and reasoning skills that are appropriate to their developmental (cognitive) age.

Intrinsic to a developmental model is an instructional process through which child and teacher work together to plan and carry out the day's activities. Because it is widely accepted that young children acquire information holistically and not as isolated and discrete units (Piaget and Inhelder, 1969), the developmental process is based upon an integrated model of learning. This model presents activities as ongoing and interrelated (Holdaway, 1985). Themes thus become extremely important vehicles for linking various skills areas. In this way, the theme of "Beginnings" might include tasks such as planting a garden or squeezing lemons to make lemonade. These tasks, in turn, might teach a range of skills such as oral language development, counting and measuring, fine-motor coordination, and social cooperation.

Instruction in the first-grade classroom under study adhered closely to the developmental framework described above. Nevertheless, what distinguished learning in this classroom from many other first-grade classrooms was the noticeable lack of workbook and worksheet activities. Paper and pencil tasks served primarily for children to record their own discoveries and to comment on experiences by means of an inventive spelling technique.[2]

---

[2] The philosophical assumption underlying the theory of inventive spelling is that children learn to write best by developing their own guidelines of spelling and grammar. In time, as these children learn more about the adult rules of the language, their child rules will be replaced by adult forms.

Journal writing was therefore an extremely important component of the curriculum; it occurred daily and encouraged children to synthesize information, to reflect critically on their experiences from multiple points of view, to relate classroom activities to one another, and to organize their ideas into a coherent text. In the *Discussion of Data* we elaborate on the relationship of the reading-writing connection to student performance outcomes.

THE THEORETICAL CONTEXT

Current research on child learning has confirmed what researchers such as Jean Piaget (1954, 1965) had proposed much earlier - that children construct their notions of the world by physically interacting with it and not by passively observing or reacting to things and events in their environment. In fact, even around age six, when youngsters acquire the capacity to think symbolically (they can use words and numbers to represent objects and relations), their thoughts are still tied to concrete reference points. As a result, even though a child of six or seven does not have to touch or move things in order to hold them in thought, his or her representation of objects and events and the relations among them still must derive from his or her own experiences with things and situations in the world. (cf. Brown, Collins, Duguid, 1989.)

We argue here that because cognitive growth in young children occurs extremely rapidly and is subject to considerable variability within generally-recognized sequences of development, it is often difficult to obtain reliable and valid results from standardized achievement tests (cf. Cohen, 1988; Meisels, 1984). Particularly when a test is structured to tap skills mastery through an abstract, decontextualized format, a mismatch occurs between the way young children naturally internalize information and the unnatural method chosen for them to communicate what they know.

If learning takes place primarily through workbooks and worksheets, students are acquiring information by rote. They may be finding out about right and wrong answers but they are not discovering how to conceptualize and analyze problems. In fact, even though students may be 'test-ready'-- that is, they may have mastered good test-taking skills, they may not have developed the theoretical constructs necessary for them to see relationships among ideas or to apply information to a variety of learning contexts. No one can deny that youngsters are excellent imitators and that imitation plays an important role in their overall physical and intellectual growth. Nevertheless, when imitation is taught at the expense of experiential learning, complex reasoning and problem-solving skills that are critical for higher-order thinking are bypassed.

Given the above arguments, many researchers (Madaus, 1983, 1988; Meisels, S.J., 1986) and child advocacy groups (National Center for Fair and Open Testing, 1987; National Association for the Education of Young Children, 1988) have advanced the contention that tests administered on a one-time basis cannot independently account for student progress or the overall success of a given educational program. They have proposed that pupil outcomes be judged in light of a range of variables that address achievement from both a qualitative and quantitative perspective. Generally, researchers concerned with the appropriate assessment of young children have proposed that interpretations of test outcomes be made in light of three factors: the degree of fit between a test and the instructional program it measures; the cultural and language preferences of students; and the intended use of the test -- that is, how results will impact a student's educational program.

METHODS

SUBJECTS

Three males and five females participated in the follow-up study of children identified as being at risk of school failure. The children reflected the cultural and socioeconomic diversity of the Boston Public Schools: four were Black, two were Hispanic, one was Asian, and one was White; five qualified for a free school lunch. At the time they were assessed on the Metropolitan Achievement Test (MAT6), the youngsters ranged in age from 75 to 85 months; four were below the age of seven and four were seven or slightly older. All subjects were judged by their teacher to be dominant speakers of English even though two heard a language other than English in the home.

As previously mentioned, subject selection was based on a child having tested poorly on the MAT6 (at or below the 40th percentile). The percentile ranks for the children in question fell between the 5th and 37th rank.

SECOND-GRADE TEACHERS OF STUDY STUDENTS

Follow-up data on the reading/language arts school performance of observed students were obtained from their second-grade teachers. Overall, these individuals had an average of 23 years of instructional experience, with 13 of these years in a second-grade classroom. Four of the teachers said they primarily used traditional instructional techniques; three said that they used a balanced blend of traditional and developmental instructional techniques; and one indicated using more developmental than traditional teaching techniques.

In February of 1989, structured telephone interviews were conducted with each of the second-grade teachers of follow-up students. The purpose of these interviews was to determine the level of success that the students were experiencing in their reading/language arts school program.

RATERS OF STUDENT WRITING SAMPLES

Because follow-up students had moved into more traditional learning environments than the one in which they had participated in first grade, we believed it would be informative to choose teachers for writing sample evaluations who were using instructional methods that articulated well with the format of the Metropolitan Achievement Test. In other words, we were looking for individuals who based a large part of their classroom instruction on workbook and worksheet learning. Two additional criteria for the selection of teachers were that: (1) they be assigned to schools having students with similar racial and socioeconomic characteristics as those of Early Learning Center students, and (2) overall, the 1987 - 1988 pupils of these teachers had attained average or better than average scores on the MAT6 Reading Test. The intent of this rater-selection process was to give us an indication as to whether the first-grade academic standards set for follow-up pupils were on a par with standards set by teachers of students learning successfully through more traditional methods.

Two teachers from different elementary schools were selected to conduct an evaluation of the writing samples of observed students. These individuals had an average of 24 years teaching experience, with 15 of these years in a first-grade classroom. The teachers were trained to rate transcripts of student journals independently. On four practice transcripts, their judgements for all items on the writing-assessment tool coincided 78% of the time. On all but one of the individual items, their judgements coincided either 75% or 100% of the time. (See Appendix A for inter-rater reliability coefficients.)

CLASSROOM OBSERVATION DATA

To determine how instruction was defined and administered, trained observers collected 12 hours of classroom data during April and May of 1988. All observations were conducted between 9:00 A.M. and 12:00 P.M. Data from visitations were quantified, and results indicated that 53% of the instructional activities focused on language development and literacy; that these activities occurred in a balanced mix of whole class, small group (fewer than 8 students), and individualized learning environments; and that nearly 40% of the activities were initiated by

the children themselves. Quantitative findings thus show that the students in question had considerable exposure to literacy activities during the school day and that a good number of these activities were defined by the children themselves.

DATA-COLLECTION INSTRUMENTS

Several indicators of student academic achievement were used to assess skills mastery. A general description of these indicators appears below with more specific descriptions appearing in the discussion of outcome data.

Metropolitan Achievement Tests (MAT6)

The Metropolitan Achievement Tests are designed to measure student academic achievement in major content areas of the school curriculum. The tests evaluate skills appearing in leading textbook series, state curriculum guidelines, and school-system system syllabuses. Scores and corresponding percentile ranks provide general information about a student's performance within a given discipline relative to the performance of other students at the same grade level. Comparison norms are established at the national level.

Pupils in the Boston Public Schools received the Reading and Math Metropolitan Tests in early May of 1988. Outcomes for only the Reading test are considered in this study.

Curriculum Referenced Tests (CRT's)

The Boston Public Schools' Curriculum-Referenced Tests (CRT's) are in-house, school-department instruments developed to articulate with the system's curriculum objectives. Each item on a CRT corresponds to an objective, such as 'listening skills,' and clusters of test items correspond to a curriculum strand such as 'phonics' or 'word recognition.' The CRT's have been highly useful to the Boston Public Schools for several reasons: First, they afford teachers an important means for monitoring the academic progress of students against a school-department criterion; second, they provide a diagnostic tool for assessing student achievement that complements the school department's norm-referenced (MAT6) testing program; and third, by systematically capturing the same kinds of data for all grade levels, the CRT's allow for student comparisons to be made across schools within the same district and at a systemwide level.[3]

Curriculum Referenced Tests were administered to the entire student body in late June of 1988. First graders received CRT's in

---

[3] The Boston Public Schools contains five school districts.

Reading/Language Arts, Writing, Math, and Science. For the purpose of this analysis, outcome data for the Reading/Language Arts and Writing CRT's will be considered.

Student Follow-up Survey

Using a telephone interview questionnaire, teachers were asked to rate specific aspects of a follow-up pupil's grammar, vocabulary, comprehension, and writing skills as compared to those of other students in the class. Teachers were also asked to evaluate a follow-up student's progress in his or her basal reading program.[4] Additionally, they were requested to indicate their instructional style from a list of styles ranging from highly developmental to highly traditional. (See Appendix B for a copy of the Student Follow-up Survey and an explanation of the rating categories comprising the instrument.

Student Writing Sample Evaluation

Earlier we explained that journal writing was an activity of primary importance in the classroom of the students under observation. This activity is one which best exemplifies a whole language approach to the teaching of literacy. Whole-language teaching integrates reading, writing, and speaking into a single cognitive framework, its purpose being to promote holistic learning by interrelating different ways of encoding meaningful relationships (cf. Holdaway, 1979). Because reading and writing are often treated as alternative means of expressing the same cognitive processes (Kucer, 1987; Wittrock, 1983), we considered writing to be a important indicator of a pupil's reading/language arts competence.

Consequently, in January of 1989, the first-grade teacher of follow-up students transcribed the texts from journals the youngsters had produced during April of 1988 -- the month prior to their being assessed in reading on the Metropolitan Achievement test. The material in the journals covered both assigned and free-choice writing. Assigned topics related to classroom projects such as the hatching of chicks, science experiments, letter-writing, and imaginary thinking. Free-choice topics generally described experiences the children had had outside of school or within their classroom. Each journal had approximately fifteen entries.

---

[4] As of the 1988 - 1989 school year, all Boston Public School's students in grades K through 8 are required to receive reading/language arts instruction from a single basal series. Comparisons of their basal reading progress are thus based upon the same instructional program.

Writing samples were transcribed exactly as they had been entered into a given journal, the context being provided for each individual entry. To facilitate the evaluation of written material by independent raters, specific transcription symbols flagged general grammatical or stylistic features of a pupil's writing that would be considered non-conventional within a school context. (See Appendix C for a summary of transcription conventions, and Appendix D for samples of transcribed texts.)

Instructions for the evaluation of student transcripts were as follows:

> *The rating categories below refer to aspects of a first-grade student's writing skill that determine his or her ability to compose clear and effective expository prose. You have seen that specific transcription symbols were used to flag non-conventional features of a student's writing. When rating each transcript, these 'non-conventional-isms' should not be considered as errors but rather as deviations from an adult standard. The adult standard used here is one that is appropriate to school settings and does not refer to standards appropriate to other dialects of English. As you know, many of the language deviations you have read in the child texts are normal for a first-grader; therefore, from a developmental or maturational point of view, they cannot be treated as 'mistakes.' Bearing this point in mind, please rate each transcript according to what you regard as 'extremely high,' 'average,' 'below average,' and 'extremely low' writing performance for a typical first grader.*

The four rating classifications mentioned above corresponded to the letters 'A,' 'B,' 'C,' and 'D,' respectively, and could be used with pluses and minuses to modify the value of any given classification. Each child's writing sample was judged according to the following criteria: (1) overall writing ability; (2) spelling and punctuation; (3) grammatical structure; (4) grammatical usage; (5) grammatical complexity; (6) descriptive modification; (7) thematic cohesion, and (8) idea development. (See Appendix E for a copy of the rating instrument.

ANALYSES OF DATA

Ratings from the teacher follow-up survey and student writing samples were entered into computer files. Pupil scores on the Metropolitan Achievement Test in Reading, and Reading/Language Arts and Writing scores for the Curriculum Referenced Tests were

also entered into computer files. The files were then analyzed using the statistical package SPSS PC+. Analyses focused on the total-group performance of subjects, student performance by age (older than seven or younger than seven) and individual student performance. Findings showed no meaningful differences for the 'age' variable and consequently, will not be discussed in this paper. Summary data were also scrutinized to locate performance differences by race and by socioeconomic status. These differences also were not found. Finally, the data were examined to determine whether second-grade teacher ratings of pupils' classroom performance were associated at all with a teacher's instructional style. The data showed no evidence of this association.

DISCUSSION OF DATA

STUDENT FOLLOW-UP SURVEY

Teacher responses to telephone questionnaire items are summarized for the total student group in **Table 1**. Categories of information parallel those used to assess children on the reading survey of the Metropolitan Achievement Test. For the general areas of grammar and composition, teachers rated a child on his or her: *vocabulary skills*, knowledge of word meanings and recognition of words in context; *phonics skills*, knowledge of beginning, medial, and final sounds; *sentence-completion skills*, an ability to choose the correct word, from a list of words, that completes the meaning for a sequence of text; *literal comprehension skills*, an ability to identify the main idea and details of a story; and *evaluative comprehension skills*, an ability to predict story outcomes and logical conclusions, and to recognize the components of a story such as the 'saddest,' 'happiest,' or 'scariest' part. Teachers judged student writing skills on the basis of two criteria: *overall grammaticality*, that is, a child's use of age-appropriate features for spelling and punctuation, grammatical structure, grammatical usage, and grammatical complexity; and *overall descriptive writing ability*, a child's literary skill in using descriptive modification, thematic cohesion, and idea development. (The reader is again encouraged to refer to Appendix E for a more elaborate description of the above evaluation criteria.)

TABLE 1

STUDENT FOLLOW-UP SURVEY
AVERAGE SKILL-AREA RATINGS

RATING SCALE: 8(A), 7(A), 6(B+), 5(B), 4(B-), 3(C+), 2(C), 1(C-)

SKILL AREA	GROUP MEAN	SCORE RANGE
VOCABULARY SKILLS	6 (B+)	4 TO 8 (B-) (A)
PHONICS SKILLS	6 (B+)	4 TO 8 (B-) (A)
SENTENCE COMPLETION SKILLS	5 (B)	3 TO 8 (C+) (A)
LITERAL COMPREHENSION	6 (B+)	3 TO 7 (C+) (A-)
EVALUATION COMPREHENSION	6 (B+)	3 TO 8 (C+) (A)
GRAMMATICAL WRITING ABILITY	5 (B)	4 TO 8 (B-) (A)
DESCRIPTIVE WRITING ABILITY	5 (B)	2 TO 8 (C) (A)

Results of teacher ratings presented in Table 1 show that on the average, students were considered to be performing at a 'B' or 'B+' level in all of the specified categories. Although scores for individual items ranged from 'C' to 'A', only one subject per category was assigned a rating lower than 'B-'.

Also comments from teachers regarding the academic performance of the students in question were all favorable. One individual who characterized her instructional style as a balanced blend of both developmental and traditional features remarked, "I really notice a difference between him and the other children. He has very good work habits and completes work on time. This may be a result of what they did with him last year. He really enjoys reading, and I know he gets support at home and that helps."[5]

---

[5] Parent involvement in their child's literacy process is an intrinsic component of the developmental education model at the Early Learning Center in question.

Another teacher with similar self-defined instructional characteristics praised a follow-up pupil's strong writing ability, explaining that "Whatever they were doing (at the ELC) in the area of writing, it was successful." Regarding three of the students in her classroom, another far more traditional teacher remarked, "I was surprised when I saw their 'MET' scores compared to the performance they are giving me. That MET test is not an indication of what a child can do!" Nevertheless, in terms of conventional school behaviors, the same teacher stated, "The kids are soooo different from any other second grader I have ever had before. They are beautiful readers, but they don't know how to sit in chairs, or they sit with their feet on chairs, or they like to sit on the floor. The three are great students...but they're into 'do your own thing.' I can't have that."

Table 2 presents a breakdown for the ratings of follow-up students' grammar and composition ability presented in Table 1. Again it should be pointed out here that the content of student-assessment categories was guided by the content of subtest areas evaluated on the Metropolitan Achievement Test. Data comparing teacher ratings of student abilities with student subtest percentile ranks for the reading survey of the (MAT6) show marked discrepancies between the assessment of a child's skills level by his or her teacher and a test's assessment of his or her skills level. For example, *vocabularly* ratings assigned by teachers are composite scores for items measuring a student's knowledge of word meanings, an ability to recognize words in context, and an ability to identify the correct word that completes the meaning for a sequence of text. If these ratings are contrasted with the percentile ranks obtained by follow-up students on the MAT6 vocabularly subtest, which also assesses the previously-mentioned skills, we notice that the information in column 1 does not correspond to the information in column 2.

In fact, with the exception of one youngster who generally scored slightly lower than the others,[6] all follow-up students were considered by their teachers to have attained a solid average or above average academic standing in grammar (vocabulary and word-recognition skills) and comprehension. The discrepancy between teacher appraisals of student classroom performance and the performance of students on the MAT6 is seen even more clearly if mean teacher ratings for specific skills areas are compared with median percentile ranks for comparable subtest areas on the MAT6.

---

[6] The teacher of this child explained that her ratings were influenced by the child's lack of motivation and not by the child's lack of training.

TABLE 2

SCORE COMPARISONS
SECOND-GRADE TEACHER RATINGS WITH MAT6 SUBTEST SCORES

RATING SCALE: 8(A), 7(A), 6(B+), 5(B), 4(B-), 3(C+), 2(C), 1(C-)

ID	VOCAB TCHR[1]	VOCAB MAT6 NAT %ILE	WORD REC[2] TCHR	WORD REC MAT6 NAT %ILE	COMP TCHR[3]	COMP MAT6 NAT %ILE
1	8 (A)	41	8 (A)	37	8 (A)	39
2	7 (A-)	22	8 (A)	47	8 (A)	46
3	7 (A-)	31	7 (A-)	37	8 (A)	28
4	4 (B-)	18	4 (B-)	42	3 (C+)	19
5	6 (B+)	18	4 (B-)	32	5 (B)	31
6	5 (B)	14	5 (B)	32	5 (B)	28
7	6 (B+)	36	8 (A)	47	5 (B)	31
8	5 (B)	10	6 (B+)	23	6 (B+)	3
	6 (B+) MEAN RATING	*20 MEDIAN %ILE*	*5 (B) MEAN RATING*	*37 MEDIAN %ILE*	*6 (B+) MEAN RATING*	*30 MEDIAN %ILE*

Note: Ratings of student performance were obtained from the second grade teacher of each target child. The content of items evaluated by teachers parallels the content of items for the appropriate subtest on the MAT6.

[1]**VOCABULARY RATINGS** ARE COMPOSITE SCORES FOR ITEMS MEASURING A STUDENT'S KNOWLEDGE OF WORD MEANINGS, ABILITY TO RECOGNIZE WORDS IN CONTEXT, AND ABILITY TO IDENTIFY THE CORRECT WORD, FROM A LIST OF WORDS THAT COMPLETES THE MEANING FOR A SEQUENCE OF TEXT.

[2]**WORD RECOGNITION RATINGS** MEASURE A STUDENT'S ABILITY TO IDENTIFY BEGINNING MEDIAL, AND FINAL SOUNDS.

[3]**COMPREHENSION RATINGS** ARE COMPOSITE SCORES FOR ITEMS MEASURING A STUDENT'S ABILITY TO IDENTIFY THE MAIN IDEA AND DETAILS OF A STORY, PREDICT STORY OUTCOMES AND LOGICAL CONCLUSIONS, AND IDENTIFY COMPONENTS OF A STORY SUCH AS THE 'SADDEST,' OR 'SCARIEST' PART.

Data in Table 3 contrast teacher appraisals of the skills level of follow-up students in second grade with test scores measuring their skills attainment at the end of first grade. A composite teacher rating for a student's overall grammar and composition ability was compared with the student's score on the school-department's Curriculum-Referenced Test in Reading/ Language Arts and with his or her percentile rank for Reading on the MAT6. The data show that the Reading/ Language Arts CRT more accurately captures a student's academic capability -- as evidenced by his or her second grade achievement -- than does the Metropolitan Achievement Test. This is because the CRT's are based directly on curriculum guides that drive the instructional content for the Boston Public Schools. Items on the Metropolitan Achievement Test, however, are drawn from a variety of curriculum sources nationally and span several grade levels,[7] which the CRT's do not do. Also of importance is that the CRT's are administered with more flexible time limits than are the Metropolitan Achievement Tests.

The data here show that when an instrument both articulates closely with a school system's curriculum and assesses student knowledge appropriate for a given grade level, that instrument provides a more realistic account of a pupil's actual and potential school performance than does a norm-referenced measurement tool covering material beyond a pupil's assigned grade. This finding raises questions about the 'curricular validity' of the Metropolitan Achievement Test for Boston Public Schools' students. More specifically, if an important goal of the MAT6 is to use objective criteria to identify students who may be in danger of failing academically, then it is doubtful whether an instrument that measures skills taught beyond a child's actual grade level can accomplish this objective with a high degree of accuracy.

---

[7] Items on the Metropolitan Achievement Test for Grade 1 test knowledge appropriate for Grades 1, 2, and 3.

TABLE 3

INDIVIDUAL SUBJECT ANALYSIS
READING LANGUAGE ARTS

RATING SCALE: 8(A), 7(A), 6(B+), 5(B), 4(B-), 3(C+), 2(C), 1(C-)

ID	GRAMMAR/COMP[1] 2ND GRADE	CRT: % CORRECT[2] 1ST GRADE	MAT6 READING[3] NAT %ILE 2ND GRADE
1	8 (A)	93	37
2	7 (A-)	80	37
3	7 (A-)	80	27
4	3 (C+)	100	20
5	7 (A-)	73	23
6	5 (B)	80	20
7	5 (B)	87	34
8	5 (B)	93	5
	6 (B+) MEAN RATING	*86 MEAN RATING*	*25 MEDIAN %ILE*

[1]**GRAMMAR AND COMPOSITION SCORE:** TEACHER RATINGS FOR VOCABULARY, PHONICS, SENTENCE-COMPLETION, LITERAL COMPREHENSION, AND EVALUATION COMPREHENSION SKILLS.

[2]**CRT READING SCORE:** PERCENT OF ITEMS SCORED CORRECTLY ON THE BOSTON PUBLIC SCHOOLS' CURRICULUM REFERENCED READING/LANGUAGE ARTS TEST. THE CRT CONTAINS ITEMS WHICH EVALUATE A STUDENT'S GRAMMAR AND COMPOSITION SKILLS.

[3]**MAT6 PERCENTILE RANKS:** BASED ON A COMPOSITE SCORE FOR THE VOCABULARY, WORD RECOGNITION, AND READING COMPREHENSION SUBTESTS.

WRITING SAMPLE EVALUATIONS

Earlier we referred to the connection between reading and writing and argued that both processes require similar cognitive operations. These operations consist of the organization of meaning through language to define, develop, classify, and conjoin ideas. Relating experience and knowledge to a text is the foundation for both reading and writing, which are active processes of composing and comprehending. More precisely stated:

> *Composing and comprehending are process oriented thinking skills which are basically interrelated. Composing...actively engages the learner in constructing meaning, in developing ideas, in relating ideas, and in expressing ideas. Comprehending...requires the learner to reconstruct the structure and meaning of ideas expressed by another writer.*
>
> (Tierney and Pearson, 1983, p. 582)

Because of this close association between reading and writing, we decided to evaluate journal material follow-up students had produced immediately prior to the assessment of their reading/language arts skills on the MAT6. We determined that a scrutiny of pupil writing samples would provide an additional measure of their reading/language arts competence at the time of testing. Results of writing evaluations by two independent raters are presented in Table 4.

TABLE 4

STUDENT JOURNAL ANALYSIS
AVERAGE SKILL-AREA RATINGS

RATING SCALE: 8(A), 7(A), 6(B+), 5(B), 4(B-), 3(C+), 2(C), 1(C-)

SKILL AREA	GROUP MEAN	SCORE RANGE
OVERALL WRITING ABILITY	6 (B+)	3 TO 7 (C+) (A-)
SPELLING AND PUNCTUATION	6 (B+)	3 TO 7 (C+) (A-)
GRAMMATICAL STRUCTURE	6 (B+)	2 TO 7 (C) (A-)
GRAMMATICAL USAGE	6 (B+)	3 TO 8 (C+) (A)
GRAMMATICAL COMPLEXITY	6 (B+)	3 TO 7 (C+) (A-)
DESCRIPTIVE MODIFICATION	6 (B+)	3 TO 8 (C+) (A)
THEMATIC COHESION	6 (B+)	3 TO 7 (C+) (A)
IDEA DEVELOPMENT	6 (B+)	3 TO 8 (C+) (A)

As expected, mean ratings of the writing samples for pupils in question showed better-than average scores for all the categories specified on the evaluation instrument. Not only did the children's texts reflect their ability to use age-appropriate grammatical features, but the texts also revealed that the children had skill in the use of descriptive terms, in relating sentences to one another, and in elaborating on a topic. Most student transcripts received scores of B+ or A-, with one transcript in each category receiving a rating of C or C+.

COMPARISONS OF FOLLOW-UP AND WRITING-SAMPLE DATA

In **Table 5** we further specify judgements of student writing ability. We begin by comparing first grade teacher ratings of a child's grammatical writing ability with second-grade teacher ratings for the same ability area. Grammar scores are composite ratings, based upon individual item ratings. The category *Grammar* refers to the use of age-appropriate spelling and punctuation, word sequences, word endings, word choice, pronoun reference, and sentence types. We also examined a pupil's stylistic writing ability. For this area, we asked first- and second-grade teachers to judge a child on his or her skill in using desscriptive modification, developing ideas, and conjoining related ideas.

Cross-comparisons of teacher judgements for both the *Grammar* and *Stylistics* categories show some similarity between contrasted ratings. Evaluations of first-grade student journals are slightly higher than second-grade teacher evaluations of overall student grammatical and writing ability. It is important to note, however, that whereas writing sample evaluations are based upon a specific product, teacher evaluations are not. Score discrepancies, then, may be attributable in part to specific versus general analyses of student performance. Nevertheless, mean ratings for the total group of observed students show a high degree of consistency across comparison categories.

An additional analysis contrasted outcome scores for follow-up students on the school-department's Curriculum-Referenced Test in Writing with ratings for student journal entries. The CRT writing test elicits three narratives, each developed around a different picture stimulus. These narratives are scored by the classroom teacher against criteria measuring the creativity of a pupil's text, its spelling, capitalization and punctuation, idea development, topic unity, and general appearance. The purpose of comparing CRT writing scores with journal scores was to contrast the judgements made by study pupils' first grade teacher with judgements of their writing ability made by other first-grade teachers.

Data show moderately high levels of agreement between teacher ratings of pupil writing competence. We have argued that writing, like reading, involves establishing relationships among words, sentences, paragraphs, and texts, and that good writing skills are closely associated with good reading skills. Teacher information regarding the progress of observed students in their second-grade basal reading program provides support for this contention. On a rating scale ranging from 'poor,' to 'excellent,' the progress of two students was judged to be 'good,' that of five students, 'very good,' and that of one student, 'excellent.' Furthermore, all the students were considered to be read-

ing at or above grade level, and some were even at the top of their class. This finding lends further credence to the argument that standardized achievement tests may not be the most appropriate measures of actual student knowledge.

TABLE 5

INDIVIDUAL SUBJECT ANALYSIS
WRITING

RATING SCALE: 8(A), 7(A), 6(B+), 5(B), 4(B-), 3(C+), 2(C), 1(C-)

ID	GRAMMAR[1] JOURNAL RATING	GRAMMAR 2ND GR TCHR	STYLISTICS[2] JOURNAL RATING	STYLISTICS 2ND GR TCHR	CRT[3] SCORE
1	6 (B+)	8 (A)	7 (A-)	8 (A)	93
2	6 (B+)	6 (B+)	8 (B+)	8 (A)	90
3	7 (A-)	6 (B+)	7 (A)	5 (B)	93
4	7 (A-)	4 (B-)	8 (A)	4 (C)	93
5	4 (B-)	5 (B)	4 (B-)	5 (B)	90
6	6 (B+)	4 (B-)	7 (A-)	4 (B-)	93
7	6 (B+)	5 (B)	6 (B+)	5 (B)	97
8	3 (C+)	5 (B)	3 (C+)	5 (B)	87
MEAN RATING	*6 (B+)*	*5 (B)*	*6 (B+)*	*5 (B)*	*92*

Note: Ratings for first-grade student journals are based upon texts written in April of 1988. The scores reported above are an average of the composite score of each of two independent raters.

[2]**GRAMMAR RATINGS OF STUDENT JOURNALS:** OBTAINED FROM ITEMS JUDGING A STUDENT'S WRITING ABILITY IN THE AREAS OF SPELLING AND PUNCTUATION, GRAMMATICAL STRUCTURE, GRAMMATIICAL USAGE, AND GRAMMATICAL COMPLEXITY.

[2]**STYLISTIC RATINGS OF STUDENT JOURNALS:** OBTAINED FROM ITEMS JUDGING A STUDENT'S WRITING ABILITY IN THE AREAS OF DESCRIPTIVE MODIFICATION, THEMATIC COHESION, AND IDEA DEVELOPMENT.

[3]**CRT WRITING SCORES:** ASSIGNED TO STUDENTS BY THEIR FIRST-GRADE TEACHER. THE SCORES MEASURE WRITING SAMPLES ON THE BASIS OF THEIR CREATIVITY, PARAGRAPH UNITY, SPELLING, CAPITALIZATION AND PUNCTUATION, AND GENERAL APPEARANCE.

CONCLUSION

In this pilot study we have looked at several measures of student achievement in the area of reading/language arts. We have shown that a standardized achievement test administered to eight first-grade pupils enrolled in a developmental education learning program did not accurately reflect either their first-grade reading/language arts skills level or their second-grade reading/language arts school performance. We have also shown that other indicators of pupil academic attainment such as teacher evaluations of student products and internally-developed school department criterion-referenced tests more accurately mirrored student academic capability and potential.

Our discussion has also centered on the interpretation of standardized achievement test results. We have explained that in the Boston Public Schools, end-of-year outcomes on the Metropolitan Achievement Test (MAT6) are used to classify students academically. A student scoring at or below the 40th percentile in a given academic area is formally labeled as being *at risk* of failing that area during the next school year. Whereas poor performance on the MAT6 may point to severe academic shortcomings for some students, this certainly was not the case for the eight *at-risk* students on whom we did follow-up work. Even though these youngsters had Metropolitan Achievement Test scores in Reading/Language Arts that placed them in the danger zone according to school-department standards, their literacy skills were at or above grade level according to teacher standards.

Finally, our pilot study raises questions about the articulation between test format and instructional format and test content and instructional content. Our findings suggest that a mismatch occurs when children acquiring skills through a concrete, hands-on, and exploratory approach to learning are evaluated on a test that taps knowledge through an abstract, decontextualized format. We have noted that young children are active learners who must construct and manipulate their environment in order to develop good conceptual skills, and that rote learning for the improvement of test performance does not promote strong critical-reasoning ability. Additionally, we have proposed that tests which assess skills across several grade levels may not accurately uncover a student's actual grade-level knowledge; consequently, these instruments may not have the curricular validity they are intended to have.

Relative to policy setting for the Boston Public Schools, the implications of our findings are both programmatic and fiscal. When the results of a single test are used to label children academically, youngsters may risk being assigned to a Chapter 1 remedial program primarily on the basis of a single criterion -- a reading or math score on the MAT6 that falls at or below the

40th percentile.[8] In the case of the eight children here, four have been receiving Chapter 1 remediation services in reading, but three of the four are among the top students in their class. At a yearly cost to the school department of almost $1300 for each child, this is may not be the wisest allocation of monetary resources.

Relative to the psychological well-being of pupils, those individuals identified as being at risk in a given subject area on the Metropolitan Achievement Test may run an even greater risk during the course of their instructional program. This risk is one of being labeled and treated as a low achiever. When teachers see an extremely low score on a standardized achievement test, they may be prone to make assumptions about a child that do not necessarily characterize his or her talents. For this reason, teachers must continually be apprised of the dangers of judging student abilities on the basis of a single measure of academic progress.

In fields such as psychology, linguistics, and anthropology, information from case studies on a few subjects have helped set the stage for research and analyses of large groups of children. Therefore, given the empirical data we have presented here, we encourage school systems that use test scores to classify children to conduct follow-up research on youngsters at the early childhood level, especially those youngsters identified as being at educational risk. In this way, research findings can inform school personnel and policy makers of the degree of fit between an assessment tool and student success in the program that tool is measuring. We also encourage school systems that use test scores to classify children to think about the limitations on the information we get from a child's test score, and with these limitations in mind, to consider relying on a variety of child performance indicators when making major educational policy decisions.

---

[8] We wish to qualify this statement by indicating that all children who are eligible for Chapter 1 services do not necessarily receive them. Placement in the Chapter 1 program depends primarily on how low the child's percentile rank is on the MAT6 as compared to other children in his/her school. Schools having a great many students with scores falling well below the 40th percentile rank will have a lower cut-off point for Chapter 1 placement than will schools having fewer students in the lowest percentile ranks.

REFERENCES

Brown, J.S., A. Collins & P. Duguid. (1989). Situated cognition and the culture of learning. 18(1), 32 - 42.

Cohen, S.A. (1988). Tests: Marked for life. New York: Scholastic, Inc.

Holdaway, D. (1979). Foundations of literacy. Portsmouth, N.H.: Heinemann Educational Books, Inc.

Kucer, S.B. (1987). The cognitive base of reading and writing. In J.R. Squire (Ed.). The dynamics of language learning (pp. 27-51). Urbana, Ill: ERIC Clearinghouse on Reading and Communication Skills.

Madaus, G.F. (1983, November). Test scores: What do they really mean in educational policy. Paper presented at the 1983 Convention of the New Jersey Education Association, Atlantic City, N.J.

Madaus, G.F. (1988). The influence of testing on the curriculum. In L.N. Tanner (Ed.). Critical issues in curriculum: Eighty-seventh yearbook of the national society for the study of education (pp. 83 - 121). Chicago: University of Chicago Press.

Meisels, S.J. (1986). Testing four- and five-year olds. Educational Leadership. 44, 90 - 92.

Medina, N. & D. Monty Neill. (1988, June) Fallout from the testing explosion: How 100 million standardized exams undermine equity and excellence in America's public schools. Cambridge, MA: National Center For Fair and Open Testing.

NAEYC position statement on developmentally appropriate practice in the primary grades, serving 5- through 8-year-olds. In S. Bredekamp (Ed.) Developmentally appropriate practice in early childhood programs serving children from birth through age 8. (pp. 64 - 84). Washington, D.C.: National Association for the Education of Young Children.

NAEYC position statement on standardized testing of young children 3 through 8 years of age. (1988, March). Young Children, 42 - 47.

Piaget, J. (1954). The construction of reality in the child. New York: Basic Books. Piaget, J. Piaget's theory. (1970). In P.H. Mussen (Ed.), Carmichael's manual of child psychology:- Vol. 1 (pp. 703-732). New York: Wiley.

Piaget, J. & Inhelder, B. (1969). The psychology of the child. New York: Basic Books.

Tierney, R.J. & P.D. Pearson. (1983). Toward a composing model of reading. Language Arts: Reading and Writing, 60(5), 568 - 580.

Wittrock, M.C. (1983). Writing and the Teaching of Reading. Language Arts: Reading and Writing, 60(5), 600 - 606.

APPENDIX A
INTER-RATER RELIABILITY COEFFICIENTS

PERCENTAGE OF INTER-RATER AGREEMENT

ITEM 1 = 75%

ITEM 2 = 50%

ITEM 3 = 75%

ITEM 4 = 100%

ITEM 5 = 75%

ITEM 6 = 75%

ITEM 7 = 100%

ITEM 8 = 75%

APPENDIX B
STUDENT FOLLOW-UP SURVEY

STUDENT FOLLOW-UP PERFORMANCE EVALUATION
EARLY LEARNING CENTER - DISTRICT A
TELEPHONE INTERVIEW PROTOCOL

(Name of Teacher), I'm going to ask you about *(Student's Name)*'s school performance in relation to the performance of other students in his/her class. I'm particularly interested in knowing how *(Student's Name)*'s skills compare to those of his/her peers in two broad content areas: Reading/Language Arts and Writing.

Using a scale of **A, B, C, D,** where **A** indicates *'extremely high performance,'* **B**, *'average performance,'* **C**, *'below average performance,'* and **D**, *'extremely low performance,'* I'd like you to judge the level of *(Student Name)*'s abilities against those of his/her classmates. You may use pluses and minuses with each letter you assign to a particular ability area. Let's begin with Reading/Language Arts. How would you rate *(Student Name)*'s:

1. **Vocabulary Skills**

   Knowledge of word meanings and recognition of words in context.

   A_____ B_____ C_____ D_____

2. **Phonics Skills**
   Identifying beginning, medial, and final sounds.

   A_____ B_____ C_____ D_____

3. **Sentence-Completion Skills.**
   Choosing the correct word, from a list of words, that will complete the meaning for a sequence of text.

   A_____ B_____ C_____ D_____

4. **Literal Comprehension Skills**
   Identifying the main idea and details of a story.

   A_____ B_____ C_____ D_____

5. **Evaluative Comprehension Skills**
   Predicting story outcomes and logical conclusions; identifying components of a story such as the 'saddest,' 'happiest,' or 'scariest' part.

   A_____ B_____ C_____ D_____

In the area of **WRITING**, how would you rate *(Student Name)*'s:

6. **Overall grammatical writing ability**

Includes age-appropriate features for: (1) **spelling and punctuation**; (2) **grammatical structure** (appropriate word sequences, the use of correct person and tense endings on verbs, the use of appropriate plural markers on nouns); (3) **grammatical usage** (appropriate word choice and pronoun reference); and (4) **grammatical complexity** (the use of different sentence types -- simple, compound, complex).

A_____ B_____ C_____ D_____

7. **Overall descriptive writing ability**

Includes: (1) **descriptive modification** (appropriate use of adjectives and adverbial modifiers); (2) **thematic cohesion** cohesion (sentences are related to each other in a meaningful way); and (3) **idea development** (the ability to elaborate on a topic).

A_____ B_____ C_____ D_____

Now I have just a few more general follow-up questions.

8. How is *(Student)* progressing in his/her basal reading program? Would you say his/her progress is:

A. excellent B. very good C. good D. fair E. poor

Name of basal reading series_____

Name of book student is using in series_____

Level of book in series_____

9. For approximately how many hours a week does *(Student Name)* engage in self-directed reading with texts other than a basal?

_____

10. For approximately how many hours a week does *(Student Name)* engage in self-directed process writing?

_____

11. In terms of a continuum of teaching styles where a **developmental teaching style** is characterized by hands-on, discovery-oriented, student-directed, and process-oriented learning and a **traditional teaching style** is characterized by paper and pencil, workbook, teacher-directed, and product-oriented learning, how would you characterize your instructional style? Would it be:

    A. completely developmental

    B. primarily developmental with some traditional features

    C. a balanced blend of developmental and traditional features

    D. primarily traditional with some developmental features

    E. completely traditional

APPENDIX C
TRANSCRIPTION CONVENTIONS

STUDENT TRANSCRIPTS
KEY TO TRANSCRIPTION CONVENTIONS

^ **OMISSION OF PUNCTUATION MARKS, CAPITAL LETTERS, OR INCORRECT PUNCTUATION**, e.g., ',' for '.'

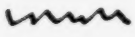
I like chicks, they are beautiful.

o **OMISSION OF A PERSON/TENSE MARKER, AN AUXILIARY VERB, OR A PLURAL MARKER ON A NOUN**

Yesterday I go to the movies; I been there many times; There are two chick in the incubator.

X **OMISSION OF A WORD OR SEQUENCE OF WORDS**

We planted and the seeds were going to grow into flowers.

 **UNCONVENTIONAL SPELLING OR RUN-ON WORDS**

I played a horsegame.

INCORRECT WORD USAGE, e.g., 'a' for 'an,' 'the' for 'a,' incorrect pronoun reference

Once upon a time there was the dog named Sam.

It's a nice horse, and I think he's pretty.

 OVERUSE OF CAPITALS/PUNCTUATION

The book is on The table near the Chair.

REVERSE WORD ORDER

He helped me it do.

(- - -) **UNCLEAR/GARBLED TEXT:** cannot decipher meaning

(=) **MEANING GLOSS:** interprets or paraphrases student's intended meaning.

APPENDIX D
SAMPLES OF TRANSCRIBED TEXTS

Context	Text	Page #
Writing about chick development incubator in room.	in The Egg There is ... a little chick^ The yolk is yellow. an^ (=And) Tamarole (=Tomorrow) The chicks will Hach^ (=hatch)	19
Free choice writing.	When I go over my cousent (=cousin's) house I will Take my Mashen (=machine) over^ in The Summer we got some Ice carem (=cream) an (=and) we got to go to toys R us^ when we got to my cousent house we rood (=rode) my bick (=bike) an (=and) we went to The store an (=and) went to The faire^ we (=we) got a new dugary (=dungaree) cote^ (=coat) we had no chraing (=training) wees^ (=wheels);	20
Free choice writing.	My los (=loose) tooth is wigaly (=wiggly) an (=and) I dont care if it fall^ the tooth firy (=fairy) will give me a dola (=dollar) or a koer^ (=quarter) she gave me a note^	23

Context	Text	Page #
Free Choice Writing.	My mom Sed (=said) She would by me a netedo (=Nintendo) viteeo (=video) game At the Ennd (=end) of spring. I am going to play a game. I am going to hook it on my TV.	18
Writing about school cookout.	We had a cook out out side in the play yard. We had a hot dog And Chips And iy-screm (=ice cream). And we playd (=played).	21
Free Choice Writing.	At the park I like to play socr (=soccer) ball outside. I am the gollee (=goalie). I catch the ball all the time.	22

Context	Text	Page #
Free choice writing - based on recent science lessons.	a comet is made out of rock and metal and Ice. @ Asteroid is made out of rock and Fire. a ring of a planit (=planet) is like a frisbe. (=frisbee) a star is wite (=white) and a star (=is) mad (=made) ou (=out) of wite rocks.	13
Writing about self as a chick.	When I was a chick I saw some horsEs eating hay in The farm. And I saw Some geeses and ducks and hens.	14
Writing about chick development - incubator in room.	In @ egg thers (=then's) a little thing. Then it grose (=grows) Bigr (=bigger) and Bigr until they hatch out from The egg.	17

APPENDIX E
WRITING SAMPLE EVALUATION INSTRUMENT

Evaluation Instrument

STUDENT WRITING SAMPLES

Transcript #_____

The rating categories below refer to aspects of a first-grade student's writing skill that determine his or her ability to compose clear and effective expository prose. You have seen that specific transcription symbols were used to flag non-conventional features of a student's writing. When rating each transcript, these 'non conventionalisms' should not be considered as errors but rather as deviations from an adult standard. The adult standard used here is one that is ap-propriate to school settings and does not refer to standards appropriate to other dialects of English. As you know, many of the language deviations you have read in the child texts are normal for a first-grader; therefore, from a developmental or maturational point of view, they cannot be treated as 'mistakes.' Bearing this point in mind, please rate each transcript according to what you regard as 'extremely high,' 'average,' 'below average,' and 'extremely low' writing performance for a typical first-grade student.

RATING SCALE
PLEASE CIRCLE THE LETTER OF YOUR RESPONSE
To qualify your judgements, you may add pluses and minuses to your letter choice, e.g., B+, C-, etc.

A__ 'extremely high'
B__ 'average'
C__ 'below average'
D__ 'extremely low'

1. *OVERALL WRITING ABILITY*

A__ B__ C__ D__

2. *SPELLING AND PUNCTUATION*
The use of age-appropriate spelling, word segmentation, and punctuation.

A__ B__ C__ D__

RATING SCALE
PLEASE CIRCLE THE LETTER OF YOUR RESPONSE
To qualify your judgements, you may add pluses and minuses to your letter choice, e.g., B+, C+, etc.

A__ 'extremely high'
B__ 'average'
C__ 'below average'
D__ 'extremely low

3. *GRAMMATICAL STRUCTURE*
Appropriate word sequences; the use of correct person and tense endings on verbs; the use of appropriate plural markers on nouns, e.g., He a real nice person; I planted a seed, and it grow into a flower. I have two best friend.

A__ B__ C__ D__

4. *GRAMMATICAL USAGE*
Age-appropriate word choice and pronoun reference, e.g. Once upon a time there was the dog named Sam; It's a nice horse, and I think he's pretty;

A__ B__ C__ D__

5. *GRAMMATICAL COMPLEXITY*
Age-appropriate use of different sentence types, e.g., **simple:** We couldn't play outside. It was raining; **compound:** It was raining and we couldn't play outside; **complex:** sentences with subordinators such as when, because, who, which, that, where, if, so, etc.: We couldn't play outside because it was raining; I know where I'm going to plant my garden; She's the witch who's good.

A__ B__ C__ D__

6. *DESCRIPTIVE MODIFICATION*

Age-appropriate use of descriptive adjectives and adverbial modifiers, e.g., Yesterday my mother bought me bright red pants and a shiny blue shirt; as compared to: My father bought me pants and a shirt.

A__ B__ C__ D__

7. *THEMATIC COHESION*
Sentences are related to each other in a meaningful way.

*A_ B_ C_ D_*

8. *IDEA DEVELOPMENT*
The student's ability to elaborate on a topic.

*A_ B_ C_ D_*